Docket No.: 0039-7484-2SRD

OIPE IAP48
SEP 28 2006
PATENT & TRADEMARK OFFICE

IN THE UNITED STATES PATENT & TRADEMARK OFFICE

IN RE APPLICATION OF: Osamu Hori et al.

SERIAL NO: 09/472,068

FILED: December 23, 1999

FOR: INFORMATION PROVIDING METHOD AND APPARATUS, AND INFORMATION RECEPTION APPARATUS

TRANSLATION OF DOCUMENT

COMMISSIONER FOR PATENTS
P.O. BOX 1450
ALEXANDRIA, VA 22313-1450

SIR:

Kenji Kobayashi, a translator residing at 2-46-10, Gokonishi, Matsudo-shi, Chiba-ken, Japan, hereby states:

(1) that I know well both the Japanese and English languages;

(2) that I translated the attached document identified as corresponding to Patent Application No. 10-372746 filed in Japan on December 28, 1998 from Japanese to English;

(3) that the attached English translation is a true and accurate translation to the best of my knowledge and belief.

DATE: September 19, 2006

BY: ____________________
Kenji Kobayashi

10 - 372746

[Name of Document] PATENT APPLICATION

[Reference Number] A009805202

[Filing Date] December 28, 1998

[To] Commissioner, Patent Office

[International Patent Classification] G06G 3/16

[Title of the Invention] IMAGE PROVIDING APPARATUS AND IMAGE PROVIDING METHOD

[Number of Claims] 8

[Inventor]

[Address or Residence] 1 Komukaitoshiba-cho, Saiwai-ku,
Kawasaki-shi, Kanagawa-ken
c/o Research & Development
Center of KABUSHIKI KAISHA
TOSHIBA

[Name] Osamu Hori

[Inventor]

[Address or Residence] 1 Komukaitoshiba-cho, Saiwai-ku,
Kawasaki-shi, Kanagawa-ken
c/o Research & Development
Center of KABUSHIKI KAISHA
TOSHIBA

[Name] Miwako Doi

[Inventor]

[Address or Residence] 1 Komukaitoshiba-cho, Saiwai-ku,
Kawasaki-shi, Kanagawa-ken
c/o Research & Development
Center of KABUSHIKI KAISHA
TOSHIBA

[Name] Kazuo Sumita

[Inventor]

[Address or Residence] 1 Komukaitoshiba-cho, Saiwai-ku,
Kawasaki-shi, Kanagawa-ken
c/o Research & Development
Center of KABUSHIKI KAISHA

	TOSHIBA
[Name]	Hideki Hirakawa
[Applicant for Patent]	
[Identification Number]	000003078
[Name]	KABUSHIKI KAISHA TOSHIBA
[Agent]	
[Identification Number]	100058479
[Patent Attorney]	
[Name]	Takehiko Suzuye
[Phone Number]	03-3502-3181
[Appointed Agent]	
[Identification Number]	100084618
[Patent Attorney]	
[Name]	Sadao Muramatsu
[Appointed Agent]	
[Identification Number]	100068814
[Patent Attorney]	
[Name]	Atsushi Tsuboi

[Appointed Agent]	
[Identification Number]	100092196
[Patent Attorney]	
[Name]	Yoshiro Hashimoto
[Appointed Agent]	

[Identification Number]	100091351	
[Patent Attorney]		
[Name]	Akira Kohno	
[Appointed Agent]		
[Identification Number]	100088683	
[Patent Attorney]		
[Name]	Makoto Nakamura	
[Appointed Agent]		
[Identification Number]	100070437	
[Patent Attorney]		
[Name]	Shoji Kawai	
[Indication of Official Fee]		
[Prepayment Register Number]	011567	
[Amount of Payment]	¥21,000.-	
[List of Items Submitted]		
[Name of Item]	Specification	1
[Name of Item]	Drawing	1
[Name of Item]	Abstract	1
[Number of General Power of Attorney]	9705037	
[Necessity of proof]	Necessary	

[Document] SPECIFICATION

[Title of the Invention] IMAGE PROVIDING APPARATUS AND IMAGE PROVIDING METHOD

[What is claimed is:]

[Claim 1] An image providing apparatus characterized by comprising:

holding means for holding images of a program;

means for classifying images of a program provided by said holding means on the basis of information obtained by means of at least one analysis method selected from moving image analysis, acoustic/speech analysis, and text analysis or on the basis of information obtained by a manual input operation, and adding an index to each of the images of the classified types so that the images can be managed in units of the classified types;

selecting means for selecting images associated with the information of the added index from the information of the added index in units of the classified types on the basis of specific information; and

associating means for obtaining the images selected in units of the classified types from said holding means, restructuring the obtained images, and providing image information.

[Claim 2] An image providing apparatus characterized by comprising:

holding means for holding images of a program;

means for classifying images of a program provided by said holding means as program images on the basis of information obtained by means of at least one analysis method selected from moving image analysis, acoustic/speech analysis, and text analysis or on the basis of information obtained by a manual input operation, and adding an index to each of the images of the classified types so that the images can be managed in units of the classified types;

selecting means for selecting images associated with the information of the added index from the information of the added index in units of the classified types on the basis of specific information;

associating means for obtaining the images selected in units of the classified types from said holding means, restructuring the obtained images, and outputting the images as image information; and

display means for displaying the output restructured image information.

[Claim 3] The image providing apparatus according to any one of claims 1 and 2, characterized in that said specific information is a keyword associated with a subject in which a user is interested.

[Claim 4] The image providing apparatus according to any one of claims 1 and 2, characterized by comprising holding means for holding images of programs to be provided, wherein said specific information is profile information registered in advance or profile information input on-line or inquiry information.

[Claim 5] The image providing apparatus according to any one of claims 1 to 4, characterized by further comprising commercial image holding means for holding commercial images, wherein said associating means obtains a predetermined commercial image from said commercial image holding means when said image is restructured, inserts the obtained image in the program so as to form a restructured image, and provides the restructured image.

[Claim 6] The image providing apparatus according to claim 4, characterized in that, when a commercial image is selected, a commercial image associated with personal interest is selected on the basis of profile information.

[Claim 7] An image providing method characterized by comprising the steps of:

constituting a database by using images of programs;

classifying images of a program provided by said database on the basis of information obtained by means of at least one analysis method selected from moving image analysis, acoustic/speech analysis, and text analysis or on the basis of information obtained by a manual input operation, and adding an index to each of the images of the classified types so that the images can be managed in units of the classified types;

selecting images associated with a keyword from the information of the added index in units of the classified types on the basis of specific information; and

obtaining the images selected in units of the classified types from said database, restructuring the obtained images, and providing the restructured image as image information.

[Claim 8] The image providing method according to claim 7, characterized by further comprising the steps of:

preparing a commercial image database for holding commercial images; and

obtaining a predetermined commercial image from said commercial image database when said image is restructured, inserting the obtained image in the program so as to form a restructured image, and providing the restructured image.

[Detailed Description of the Invention]

[0001]

[Technical Field of the Invention]

The present invention relates to an image providing apparatus and an image providing method for selecting an image of user's interest from a large number of programs supplied from program providers and providing the user with the selected program.

[0002]

[Prior Art]

In recent years, growth of information infrastructures are boosting opportunities for distributing to homes many digital images through CATV (cable television broadcasting), digital satellite broadcasting, or the Internet. In these media, a variety of programs are provided, and the number of service channels has reached a number of several hundreds or several thousands.

[0003]

Therefore, it is becoming difficult for a user to appropriately select a program from the several hundred or several thousand channels or several tens of thousands or more programs included in the channels.

[0004]

To solve this problem, a receiver device for automatically recording programs of a user's interest using information of an electronic program list sent from a broadcasting station has been proposed (e.g., "video device" disclosed in Jpn. Pat. Appln. KOKAI Publication No. 7-135621).

This proposed device has a function of selecting programs in which a user may be most interested from the information of a program list on the basis of keywords registered in advance. Even with this device, however, the selection can be roughly performed only in units of programs.

[0005]

In a program such as a news show and a variety show, one program is constituted in units such as "topics" and "corners". In many cases, a user is only interested in some images in one program.

[0006]

However, in automatic recording in units of programs, one program is entirely selected and recorded from the beginning to the end. The user must watch the recorded program from the beginning to the end.

[0007]

[Objects of the Invention]

A system for automatically selecting a program in accordance with the user's requirement from an enormous number of programs provided by a program provider is proposed. In this system, however, a program in which a user may be interested is selected in units of programs by using the information of an electronic program list sent from the broadcasting station.

[0008]

In the above system, a program in which a user may be most interested is selected from the information of a program list on the basis of keywords registered in advance. However, since the selection is performed in units of programs, the above system poses many problems.

[0009]

Consider a program such as a news show or a variety show. In such programs, one program is constituted of units of "topics" or "corners". Quite often, the user is only interested in some images in one program.

[0010]

However, in automatic recording in units of programs, one program is entirely selected and recorded from the beginning to the end. The user cannot know the position of the information of his/her actual interest unless he/she watches the entire program.

[0011]

Hence, even when a program is selected and recorded by filtering, the user must watch the recorded program from the beginning to the end, wasting the recording medium and the user's time.

[0012]

Accordingly, an image processing apparatus capable of reliably selecting only a part of a program in which the user is actually interested from a large number of broadcast programs, not by filtering in units of programs, is ardently required.

[0013]

Accordingly, it is a first object of the present invention to provide an image providing apparatus and an image providing method capable of appropriately selecting and recording only portions of user's actual interest from a large number of broadcast programs, not by filtering in units of programs.

[0014]

It is a second object of the present invention to provide an image providing apparatus and an image providing method capable of solving a further problem that, when only portions of a user's actual interest are appropriately selected, commercial messages which are required by the program provider side to be watched and listened to are omitted by the selection.

[0015]

[Means for Achieving the Object]

In order to achieve the above-mentioned objects, an image providing apparatus according to the present invention is characterized by comprising: holding means for holding images of a program; means for classifying images of a program provided by the holding means on the basis of information obtained by means of at least one analysis method selected from moving image analysis, acoustic/speech analysis, and text analysis or on the basis of information obtained by a manual input operation, and adding an index to each of the images of the classified types so that the images can be managed in units of the classified types; selecting means for selecting images associated with the information of the added index from the information of the added index in units of the classified types on the basis of specific information; and associating means for obtaining the images selected in units of the classified types from the holding means, restructuring the obtained images, and providing image information.

[0016]

In this system, images of programs are formed into a database, and images provided by the database are classified on the basis of information obtained by means of at least one analysis method selected from moving image analysis, acoustic/speech analysis, and text analysis or on the basis of information obtained by a manual input operation, into scene units, window units, or the like. An index is added to each of the images of the classified types so that the images can be managed in units of the classified types. Images associated with the information of the added index are selected from the information of the added index in units of the classified types using specific information as a keyword. The images selected in units of the classified types are obtained from the database, restructured, and provided as image information.

[0017]

More specifically, information representing the contents of video data is obtained by means of moving image analysis means, acoustic/speech analysis means, text analysis means, or manual input means, for various video data provided as programs. The video data is classified in detail on the basis of the information so that the video data can be managed in units of the classified types. The apparatus includes means for adding an index (tag) to each of the images of the classified types. Corresponding partial video data is selected from one or a plurality of video data in the index (tag) information on the basis of specific information, such as personal profile data registered in advance, profile data input on-line, and inquiry information. The selected images are associated with the index information so as to be displayed and easily watched, thereby providing a user with only an image necessary for the user.

[0018]

In the case where a broadcast program is not a chargeable program but a no-charge program taking advertisement rates as a source of revenue, getting a high audience rating is an important factor for getting a program providing sponsor. In this case, in order to solve the problem that commercial messages are not the object of the audience rating, necessary commercial messages are selected from a commercial message bank prepared by the program provider and inserted in a restructured image.

[0019]

All of these items of processing are too heavy for a single receiver device owned by a user to process. In order to solve the problem, there is provided a system having a client/server type system structure in which video analysis is performed on the program provider side or in a relay point, and restructuring of an image is performed in the receiver device of the user. Alternatively, restructuring of an image may be committed to the server side from the client side, and a function of displaying only a result may be imparted to the client side.

[0020]

As a result of this, a problem that a program, an image of which is only required to be partially watched and listened to is entirely made to be an object of recording or watching and listening to, can be solved by making it possible to manage the images in units of the classified image types. Furthermore, in the case where only a portion of an image is cut out therefrom so as to be made to be an object of recording or watching and listening to, a problem that commercial messages which are required by the program provider side to be watched and listened to are omitted, can also be solved. When such a system is constituted, the system is required to execute a heavy processing of video analysis on the one hand, and is required to perform selection of a program and association for a user on the other hand. A system which can meet the requirement of load dispersal and individual correspondence can be provided.

[0021]

[Embodiment of the Invention]

An embodiment of the present invention will be described below with reference to the accompanying drawings.

[0022]

(Basic structure of the present invention)

An embodiment is shown in a block diagram of FIG. 1 as a basic arrangement of the present invention. In FIG. 1, numeral 101 denotes a digital image database for storing images of the programs as digital data. Numeral 102 denotes a video analysis section, which subjects each of the images stored in the digital image database to a predetermined analysis process for each program so as to obtain a database of the video analysis result 103. The video analysis result of the video analysis section 102 is managed in a database which is randomly accessible.

[0023]

Numeral 104 denotes a user profile database, numeral 105 denotes an image selection section, 106 denotes a link section, and 107 denotes a display section. Of these constituent components, the user profile database 104 is a file in which information on a user's own taste or information on a field in which the user is interested is registered, and is managed in a database in units of users.

[0024]

The image selection section 105 searches the database storing the video analysis result 103 of the video analysis section 102 for data meeting the information on the user, thereby selecting a portion of an image in which the user is interested. The link section 106 reads out the corresponding image portion selected by the image selection section 105 from the digital image database 101 so as to restructure the image, and outputs the image. The display section 107 is a display device provided on the user side, for displaying the image read out by the link section 106 from the digital image database 101 and restructured by the section 106.

[0025]

In this system having the arrangement described above, the object of service is, for example, a program provided on the broadcasting station side or the client side of the Internet, and the image data which are the object of processing are a plurality of programs. The programs whose images are subjected in advance to analog-to-digital conversion are used in the system, and stored and managed in the digital image database 101.

[0026]

The digital data may be MPEG-2 compressed data or DV compressed data.

[0027]

The digital images have "title names" in units of programs and "frame numbers" in units of frames in each program and are stored in a medium, e.g., a hard disk which can be accessed from an arbitrary position.

[0028]

The medium is not limited to the hard disk and may be another new type of medium such as a DVD-RAM (ROM) capable of being randomly accessed. In other words, the medium is only required to be accessed at a desired section thereof by designating a "title" and a "frame number". The digital image data need not maintain the image size and quality of the original analog data. A compression scheme such as MPEG-1 or MPEG-4 that saves the image capacity may be employed depending on the adopted application.

[0029]

In this system, output from the digital image database 101 storing such video data is supplied to the video analysis section 102, which subjects each of the programs to analysis processing and stores information of the video analysis result 103. The storage destination is the video analysis result database, which is a database which can be randomly accessed. The information of the video analysis result 103 is managed in this database.

[0030]

As for video analysis utilized in the video analysis section 102, a technique of determining the video data structure on the basis of information of a cut with an instantaneous change in a video scene or camera movement (pan or zoom) using moving image analysis means that has conventionally been studied, and adding an index (tag) to an image is used.

[0031]

The position where the scene instantaneously changes can be detected by comparing similar frame images with each other. The similarity can be obtained by calculating the histogram of the frequency of a color in each image and comparing the histograms with each other. A portion with low similarity is a point where the scene instantaneously changes.

[0032]

To obtain a camera movement parameter, optical flows representing the positions of movement of pixels are obtained from two images. Assuming that most optical flows are obtained from the background, the movement of the camera is calculated on the basis of dominant optical flows.

[0033]

When the camera is panning, most optical flows appear parallel to each other. When the camera is zooming, optical flows are directed to a certain point. Details are described in reference (1), Hirotada Ueno, Takafumi Miyabu, and Satoshi Yoshizawa, "Proposal of Interactive Video Editing Scheme Using Recognition Technology", IECE Papers (D-II), VOL. J75-D-II, No. 2, pp. 216 - 225 and reference (2), Masahiro Shibata, "Video Contents Description Model and Its Application to Video Structuring", IECE Papers (D-II), VOL. J78-D-II, No. 2, pp. 754 - 764.

[0034]

In the acoustic/speech analysis means, music and the human voice can be separated from each other because music has few mute portions and frequency components that cannot be found in the human voice, and the human voice can be discriminated because it has characteristic features reverse to those of music, and the male voice and the female voice have a pitch difference.

[0035]

Details of the method of discriminating between the male voice and the female voice are described in reference (3), Keiichi Minami, Akihito Akutsu, Hiroshi Hamada, and Yoshinobu Sotomura, "Video Indexing Using Sound Information and Its Application", IECE Papers (D-II), VOL. J81-D-II, No. 3, pp. 529 - 537, and a detailed description thereof will be omitted.

[0036]

With this method, an index can be added to an image on the basis of video information and speech information.

[0037]

For example, sound data is analyzed to separate a music portion from a portion of male/female voice. Then, video scenes associated with the sound data are sorted into scenes associated with the music portion, scenes associated with the male voice, and scenes associated with the female voice, and indexes are added to the respective scenes.

[0038]

If character data associated with video data accompanies the video data, the text is analyzed to be used in index addition. In the U.S.A., video data contains character data called "closed caption". If such data can be used, text analysis using the conventional natural language processing technology can be performed to perform indexing according to the contents.

[0039]

That is, on the basis of character data accompanying an image, an index based on the analysis result of character data contents can be added to the image.

[0040]

In addition to the index addition performed by automatic index addition based on the various analysis as described above, index addition can also be performed by manual operation at need. If automatic index addition and manual index addition are used simultaneously, more exact index addition can be realized.

[0041]

On the other hand, on the client side, the user profile database 104 in which information on the taste or the field of interest of each user is registered is prepared for providing the user with a service. The user profile is prepared by inquiring of the user or obtaining information in advance through a questionnaire. The user profile has text information having information including keywords representing the taste of a user.

[0042]

Key words include various key words such as a "name of a favorite movie actor", a "name of a favorite sports player", a "humorous conversation", "golf and the game of GO" which are the users interests.

[0043]

The image selection section 105 searches for partial video data meeting the user's taste on the basis of information of the user profile database 104 and information of the video analysis result 103. To search for the partial video data, it is only required to find out video data matching the keyword.

[0044]

In this search for an image matching the keyword on the basis of the video analysis result 103, an image matching keywords similar to the user profile can also be detected using a thesaurus (dictionary of synonyms or taxonomy, or index for information search), thereby searching for corresponding partial video data.

[0045]

With the image selection section 105, associated video data can be finely identified/searched for in units of scenes, units associated with speech data, units associated with character data or the like, so that partial video data of each user's interest can be selected and extracted.

[0046]

The search result thus obtained is supplied to the link section 106. The link section 106 reads out the corresponding image portion from the digital image database 101 on the basis of the selection result, restructures the video data, and sends the restructured video data to the display section 107 of the user. The display section 107 displays the restructured video data. The user can watch and listen to partial video data which has no useless part and in which he/she is interested.

[0047]

An outline of the basic arrangement of this system has been described above.

[0048]

Methods of implementing the individual processing will be described below in detail.

[0049]

(Application to server/client type system)

In order to apply the above system to a server/client type system, an arrangement shown in FIG. 2 or 3 need only be considered. More specifically, in the arrangement shown in FIG. 2, the digital image database 101, video analysis section 102, image selection section 105, and link section 106 are provided on the server side, and the display section 107 and user profile database 104 are provided on the client side.

[0050]

In the arrangement shown in FIG. 3, the digital image database 101, video analysis section 102, and image selection section 105 are provided on the server side, and the link section 106, display section 107, and user profile database 104 are provided on the client side.

[0051]

As described above, when the system is structured as a server/client type system, an arrangement may be considered in which only the section for preparing the user profile and sending it to the server on-line, and the section for receiving and displaying the result are arranged on the client side as shown in FIG. 2. Alternatively, an arrangement in which a section for performing association with video information on the basis of the selection result, and restructuring the video data into a form of display arranged on the client side, may be employed.

[0052]

Allotment of the functions to the client side is determined depending on the processing capability of the client side.

[0053]

Furthermore, an arrangement in which the digital image database 101 and video analysis section 102 are provided on the server side, and the image selection section 105, link section 106, display section 107, and user profile database 104 are provided on the client side as shown in FIG. 4, may also be considered.

[0054]

As shown in the arrangement of FIG. 4, when the sections other than the video analysis section are allotted to the client side, it is necessary to download the processing result to the client side. Accordingly, this arrangement depends not only on the processing capability of the client side but also on the information storage capability, and the line capability for downloading.

[0055]

However, since the above system arrangement has an effect of processing distribution, it is an effective arrangement when the client side has high capability and the line is that of CATV, optical fiber, or intranet.

(Details of video analysis section)

Next, details of processing performed by the video analysis section 102 will be described below.

[0056]

FIG. 5 shows the flow of processing so as to explain details of an example of processing performed by the video analysis section 102. The video analysis section 102 can analyze all video data stored in the digital image database 101 so as to obtain a video analysis result 103. In this case, all the video data are analyzed for each program.

[0057]

Video data contains not only data of an image but also sound and text data. Hence, analysis of video data is performed in three steps: text analysis, moving image analysis, and acoustic/speech analysis. The processing order is not particularly specified.

[0058]

As for text analysis, closed caption information in the video data is extracted (steps S1 and S2), morphemes are analyzed (step S3), and keywords are analyzed on the basis of the morpheme analysis result (step S4). The above analysis is performed for all video programs in the image database.

[0059]

As for moving image analysis, a cut of a moving image in video data is detected (steps S1 and S5), the camera movement parameter is extracted (step S6), and the video data is segmented on the basis of the camera movement parameter (step S7). This analysis is performed for all video programs in the image database.

[0060]

As for acoustic/speech analysis, acoustic identification is performed in video data (steps S1 and S8), speech recognition is performed (step S9), and keywords are extracted on the basis of the recognition result (step S10). This analysis is performed for all video programs in the image database.

[0061]

Text analysis, moving image analysis, and acoustic/ speech analysis are completed, and produce analysis results.

[0062]

By video analysis according to these procedures, various pieces of information on the video data are obtained. The pieces of information are processed by high-level integration processing (step S11) for integrating the individual information.

[0063]

As for text analysis, moving image analysis, and acoustic/speech analysis, conventionally known analysis techniques can be used as has already been described above.

[0064]

For example, in text analysis, a closed caption contained in video data is extracted, and the roles of words are analyzed by morpheme analysis. An important keyword describing a scene such as a proper noun is extracted from the words. As the keyword, not only a proper noun but also information representing a high frequency of occurrence is also used.

[0065]

In moving image analysis, video data is segmented by extracting a scene with an abrupt change or camera movement information (reference (1)). In acoustic/speech analysis, music data and speech data are separated from each other by acoustic identification, male voice and female voice are separated from each other by speech recognition (reference (3)), and a keyword is extracted by using speech recognition.

[0066]

Integration processing aims at storing information obtained by the individual processing as a database, in association with each other, and integrating the information to generate a new keyword.

[0067]

For example, a process of associating individual processing operations with each other is performed in the following way.
Assume that processing is to be performed in units of segmented video data, and a keyword is present as an important proper noun in the video data. Even when the keyword is obtained from the caption (comment or explanation), video frames corresponding to the position of the keyword cannot be accurately known.

[0068]

The position of the keyword is identified using speech recognition, and the keyword is added to partial video data at a position with consecutive speech data.

[0069]

The analysis result is generated as, for example, a table as shown in FIG. 6. In FIG. 6, the title of the program is "news", keywords representing the characters and situation are "politics", "economy", and "weather forecast", and "0:00 - 0:05", "0:15 - 0:16", and "0:23 - 0:25" are picked up as window appearance times (frames) associated with the respective keywords. That is, video data is segmented in reference to time (frames) in units of program titles, and important keywords appearing in the frames are added to form a table.

(Details of processing performed by image selection section)

Next, details of processing performed by the image selection section 105 will be described below.

FIG. 7 shows a flowchart showing the processing flow describing the processing performed by the image selection section 105 in detail.

Description will be given with reference to FIG. 7. The image selection section 105 refers to information in the database of the video analysis result 103 and user profile database 104 to search for partial video data of interest to the user.

[0070]

Keywords are selected from the user profile database 104 one by one, and associated words are picked up using the thesaurus dictionary (steps S21 and S22).

[0071]

After picking up the associated words, the picked up associated words are collated with words represented in the video analysis result. If an associated word and a word in the video analysis result match each other, information representing the position of the partial video data and the title to which the frame belongs is recorded (steps S23, S24, and S25). In keyword matching, if the same associated word recurs, it is made an object of matching.

[0072]

Processing performed by the image selection section 105 has been described above in detail.

(Information of partial video data acquired by keyword matching)

FIG. 8 shows an information example of partial video data acquired by keyword matching. In this case, one keyword in the user profile database 104 is "animal". Information on animal is searched for using thesaurus data, and leading characters such as "horse" and "ox" are selected and collated with keywords in the database of the video analysis result 103 to obtain information of corresponding partial video data. The information of the obtained corresponding partial video data is recorded so as to obtain a record of the result of keyword matching as shown in FIG. 8.

(Link section)

Next, details of the link section 106 will be described below.

FIG. 9 shows views for explaining the processing performed by the link section 106. The link section 106 obtains information from the image selection section 105, obtains partial video data from the digital image database 101 at need, and performs video data association processing.

[0073]

The link section 106 not only prepares associated information for the meaning, but also performs processing for constituting a window for displaying the information for the user. There is a method in which association processing for the meaning and structuring a window for displaying information for the user are separately performed. However, in this case, in order to simultaneously perform the above two processing operations, association processing and window structuring are performed by using the HTML (hypertext markup language used in Web).

[0074]

First, it is checked whether or not processing for all the keywords is ended. If processing is not ended, processing is continued (step S31). Partial video data selected in association with one keyword is acquired from the digital image database 101 (step S32). To acquire partial video data from the digital image database 101 by random access at a sufficiently high speed, time code information (frame information) can be directly used as it is.

[0075]

Otherwise, a copy of partial video data, partial video data with a reduced window size, or a copy of partial video data using a different compression ratio or compression scheme is acquired.

[0076]

One or a plurality of still image frames of acquired partial video data are acquired (step S33), and used as materials for constituting the window. The keyword is associated with the still image frame, and the still image frame is associated with the partial video data (steps S34 and S35). Information of the still image frame is described using the HTML (step S36).

[0077]

When partial video data selected in accordance with a keyword is processed, the next keyword is processed. Otherwise, the above processing is repeated (step S37).

[0078]

It is determined whether or not processing for all keywords is ended (step S31). If processing for all the keywords is ended, the contents described by the HTML are output or sent to the display section (step S38).

Otherwise, processing is continued.

[0079]

FIG. 10 shows an example of a window generated by the link section 106 in the above-mentioned manner. In this example, keywords in the user profile are "shopping", "public facilities", "transportation/bank", and "health/hospital". Therefore, partial video data of programs associated with words such as "department store" and "bakery" associated with "shopping" are acquired, and still image frames each of which is one frame of partial video data are pasted in line like indices. CMs arranged sporadically are advertisements of sponsors.

[0080]

In the window shown in FIG. 10, each still image frame is linked to corresponding partial video data such that the partial video data is displayed by a click button.

[0081]

In order to generate such a window, a necessary description need only be prepared using the HTML. HTML is an abbreviation for HyperText Markup Language, which means a page description language used as the general format of information provided by the WWW or W3 (World Wide Web) service of the Internet. HTML is based on SGML (Standard Generalized Markup Language) and can designate the logical structure of a document and link between documents by inserting a markup called a "TAG" in the document.

[0082]

WWW is a client/server information service on the Internet. A network user can access information using a dedicated Web browser. Provided information are HTML documents called home pages, Web pages, or WWW pages connected by the hyper link. Information can be displayed by tracking the link.

[0083]

Documents handled by WWW can include multimedia information, and the server side can execute a program to perform special processing. This function can be used to provide a unique information search service.

[0084]

(CM insertion processing)

Next, a method of inserting advertisements of sponsors, i.e., CMs in the partial video data will be described below.

[0085]

A CM of a still image such as a motion GIF may be pasted in the frame. Alternatively, a commercial film may be inserted in an appropriate portion of partial video data. A CM associated with a keyword is selected from a CM bank storing CMs, and the selected CM is inserted in the partial video data, whereby a CM of the user's interest can effectively attract the user's attention. Alternatively, CMs may be inserted in the partial video data regardless of the user's taste.

[0086]

As for the method of selecting CMs, a tag can be added to a CM since CM is also video information as described above. Alternatively, information on the CM may be manually input in advance.

[0087]

Information on the CM is searched for on the basis of a keyword in the user profile database 104, thereby selecting a CM of highest relevance.

[0088]

Details of the present invention have been described above. In short, an image providing apparatus according to the present invention is characterized by comprising: holding means for holding video data of a program; means for classifying images of a program provided by the holding means on the basis of information obtained by means of at least one analysis method selected from moving image analysis, acoustic/speech analysis, and text analysis or on the basis of information obtained by a manual input operation, and adding an index to each of the images of the classified types so that the video data can be managed in units of the classified types; selecting means for selecting video data associated with the information of the added index from the information of the added index in units of the classified types on the basis of specific information; and associating means for obtaining the video data selected in units of the classified types from the holding means, restructuring the obtained video data, and providing video information.

[0089]

In this system, video data of programs are formed into a database, and video data provided by the database are classified on the basis of information obtained by means of at least one analysis method selected from moving image analysis, acoustic/speech analysis, and text analysis or on the basis of information obtained by a manual input operation, into scene units, window units, or the like. An index is added to each of the images of the classified types so that the video data can be managed in units of the classified types. Images associated with the information of the added index are selected from the information of the added index in units of the classified types using specific information as a keyword. The video data selected in units of the classified types are obtained from the database, restructured, and provided as video information.

[0090]

More specifically, information representing the contents of video data is obtained by means of moving image analysis means, acoustic/speech analysis means, text analysis means, or manual input means, for various video data provided as programs. The video data is classified in detail on the basis of the information so that the video data can be managed in units of the classified types. The apparatus includes means for adding an index (tag) to each of the images of the classified types. Corresponding partial video data is selected from one or a plurality of video data in the index (tag) information on the basis of specific information, such as personal profile data registered in advance, profile data input on-line, and inquiry information. The selected video data are associated with the index information so as to be displayed and easily watched, thereby providing a user with only video data necessary for the user.

[0091]

In the case where a broadcast program is not a chargeable program but a no-charge program taking advertisement rates as a source of income, getting a high audience rating is an important factor for getting a program providing sponsor. In this case, in order to solve the problem that commercial messages are not the object of the audience rating, necessary commercial messages are selected from a commercial message bank prepared by the program provider and inserted in the restructured video data.

[0092]

All of these items of processing are too heavy for a single receiver device owned by a user to process. In order to solve the problem, there is provided a system having a client/server type system structure in which video analysis is performed on the program provider side or in a relay point, and restructuring of video data is performed in the receiver device of the user. Alternatively, restructuring of video data may be committed to the server side from the client side, and a function of displaying only a result may be imparted to the client side.

[0093]

As a result of this, a problem that a program, video data of which is only required to be partially watched and listened to is entirely made to be an object of recording or watching and listening to, can be solved by making it possible to manage the video data in units of the classified image types. Furthermore, in the case where only a portion of video data is cut out therefrom so as to be made to be an object of recording or watching and listening to, a problem that commercial messages which are required by the program provider side to be watched and listened to are omitted, can also be solved. When such a system is constituted, the system is required to execute a heavy processing of video analysis on the one hand, and is required to perform selection of a program and association for a user on the other hand. A system which can meet the requirement of load dispersal and individual correspondence can be provided.

[0094]

Further, the present invention is not limited to the embodiment described above, and can be variously modified to be implemented.

[0095]

[Advantage of the Invention]

As has been described above, according to the present invention, only video data of portions which are actually required by the user who is watching the program can be recorded or reproduced without recording or reproducing the entire program. In addition, partial video data (video data in units of types) are associated with each other and restructured so as to result in a visually convenient display. Furthermore, the problem that commercial messages that the program provider wants the viewer to watch are omitted when only part of video data is selected and recorded or watched, can be solved.

[Brief Description of the Drawings]

[FIG. 1]

FIG. 1 is a view for explaining the present invention and is a block diagram showing an example of the basic arrangement of an image providing apparatus according to the present invention.

[FIG. 2]

FIG. 2 is a view for explaining the present invention and is a block diagram showing a specific arrangement example of the image providing apparatus.

[FIG. 3]

FIG. 3 is a view for explaining the present invention and is a block diagram showing another specific arrangement example of the image providing apparatus.

[FIG. 4]

FIG. 4 is a view for explaining the present invention and is a block diagram showing still another specific arrangement example of the image providing apparatus.

[FIG. 5]

FIG. 5 shows views for explaining the present invention and shows a flowchart showing a processing flow of an example of processing performed by a video analysis section 102.

[FIG. 6]

FIG. 6 is a view for explaining the present invention and is a view showing a table representing an example of an analysis result obtained by the video analysis section 102.

[FIG. 7]

FIG. 7 shows views for explaining the present invention and shows a flowchart showing a processing flow of an example of processing performed by an image selection section 105.

[FIG. 8]

FIG. 8 is a view for explaining the present invention and is a view showing information examples of partial video data collected by keyword matching performed by the image selection section 105.

[FIG. 9]

FIG. 9 shows views for explaining the present invention and shows a flowchart for explaining the processing performed by a link section 106.

[FIG. 10]

FIG. 10 is a view for explaining the present invention and is a view showing examples of frames prepared by the link section 106.

[Explanation of Reference Symbols]

101 ... Digital image database,

102 ... Video analysis section,

103 ... Video analysis result,

104 ... User profile database,

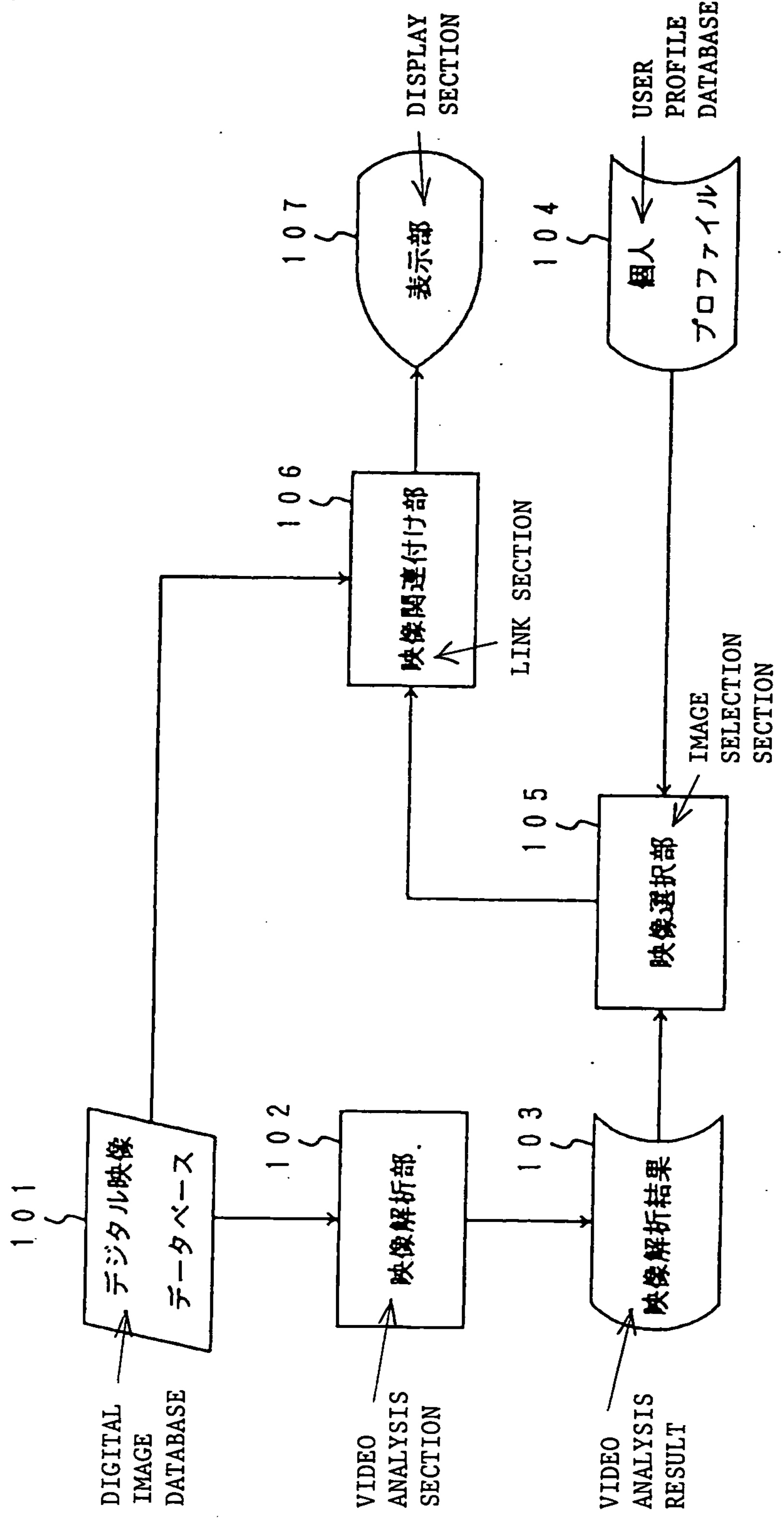
105 ... Image selection section,

106 ... Link section,

107 ... Display section.

SEP 28 2006
PATENT & TRADEMARK OFFICE

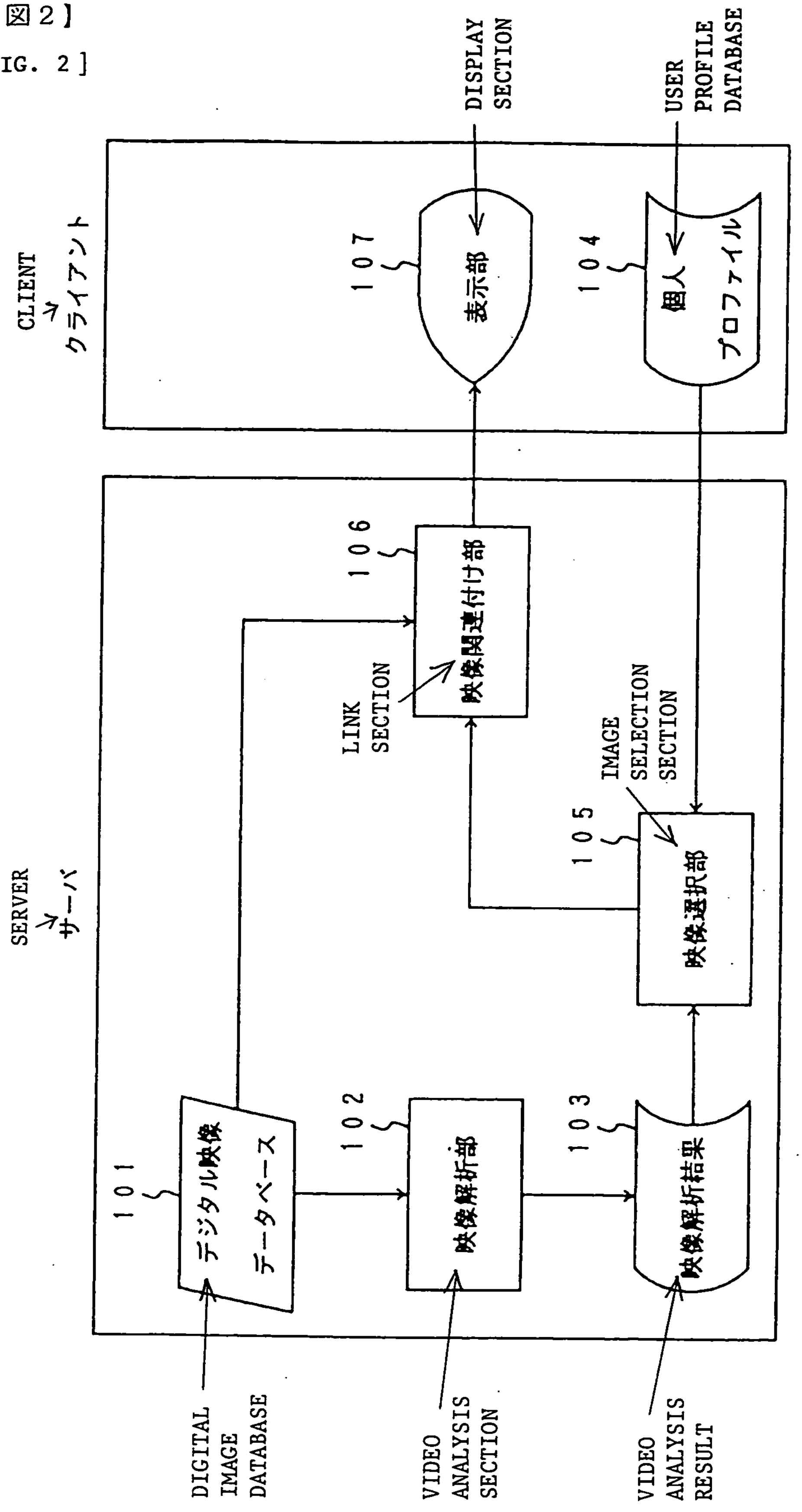
【書類名】　図面
NAME OF DOCUMENTS　DRAWINGS
【図1】
[FIG. 1]


101
デジタル映像データベース
DIGITAL IMAGE DATABASE
102
映像解析部
VIDEO ANALYSIS SECTION
103
映像解析結果
VIDEO ANALYSIS RESULT
104
個人プロファイル
USER PROFILE DATABASE
105
映像選択部
IMAGE SELECTION SECTION
106
映像関連付け部
LINK SECTION
107
表示部
DISPLAY SECTION

処理の基本構成の一実施例を示す図
VIEW SHOWING AN EXAMPLE OF BASIC ARRANGEMENT OF PROCESSING

【図２】

[FIG. 2]



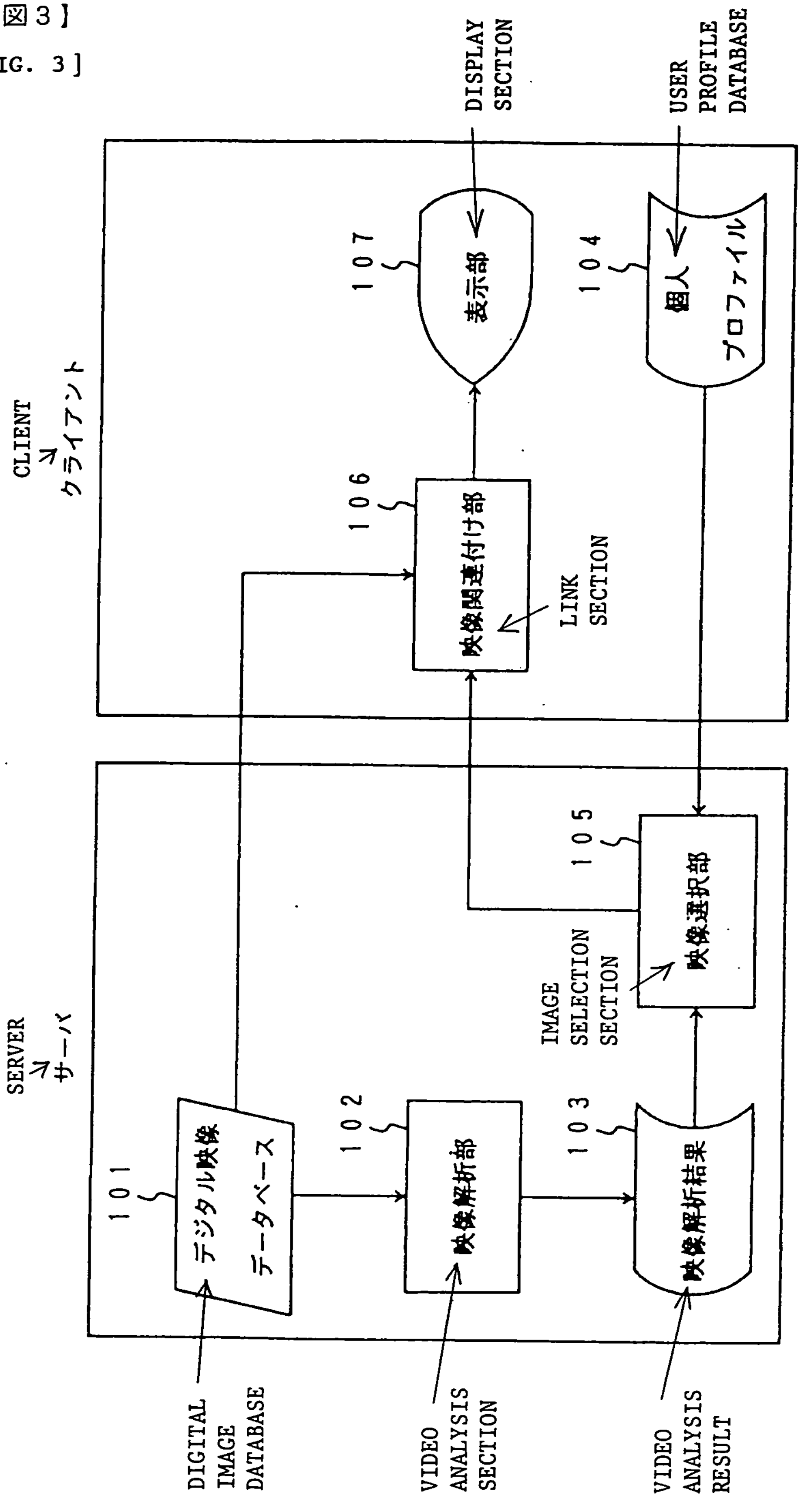
CLIENT
クライアント
107
表示部
DISPLAY SECTION
104
個人プロファイル
USER PROFILE DATABASE
SERVER
サーバ
101
デジタル映像データベース
DIGITAL IMAGE DATABASE
102
映像解析部
VIDEO ANALYSIS SECTION
103
映像解析結果
VIDEO ANALYSIS RESULT
105
映像選択部
IMAGE SELECTION SECTION
106
映像関連付け部
LINK SECTION

サーバ／クライアント処理の一構成(1)

AN EXAMPLE OF ARRANGEMENT OF SERVER/CLIENT PROCESSING (1)

【図３】

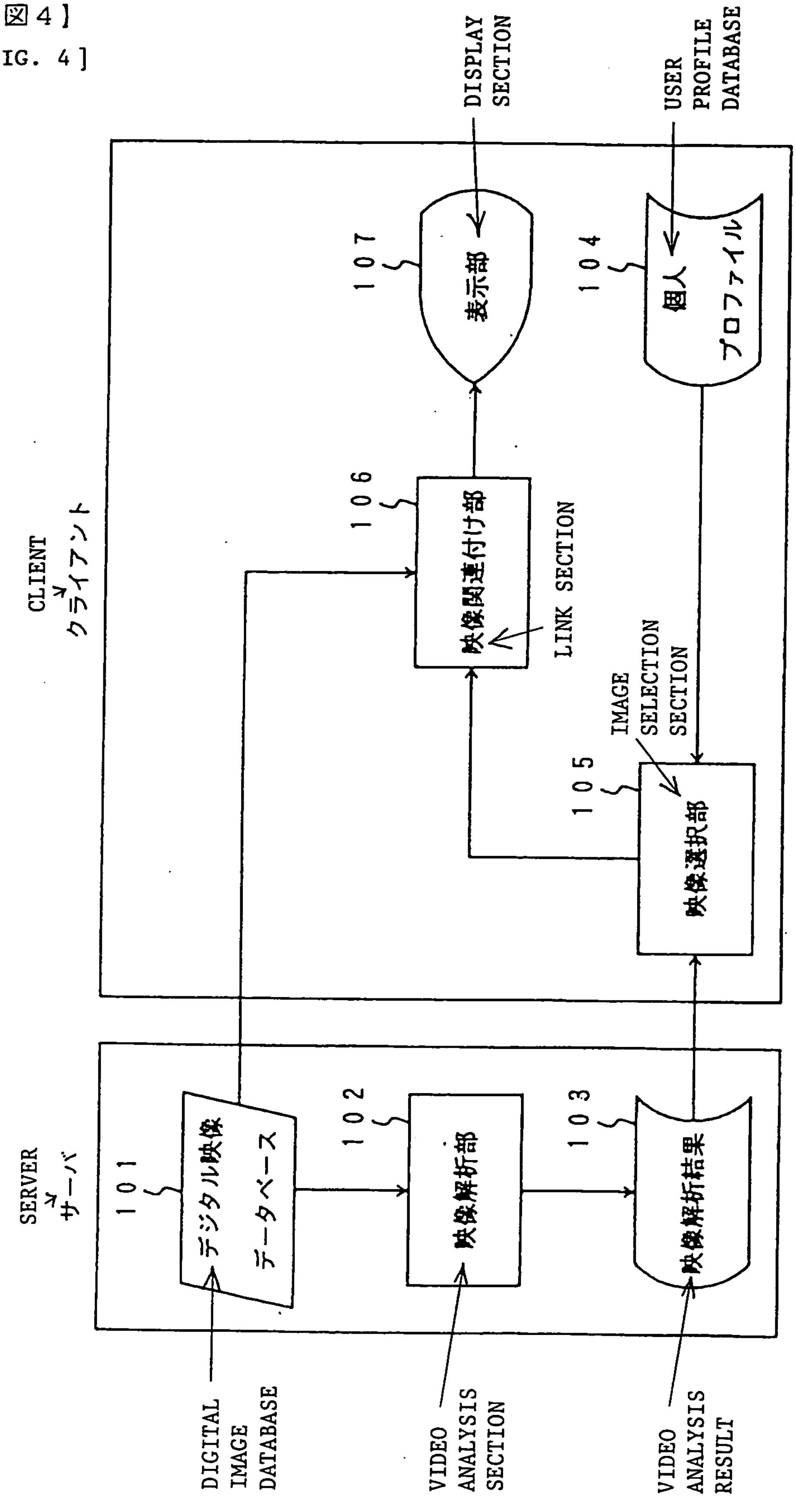
［FIG. 3］


SERVER
サーバ
CLIENT
クライアント
101
デジタル映像データベース
102
映像解析部
103
映像解析結果
104
個人プロファイル
105
映像選択部
106
映像関連付け部
107
表示部
DIGITAL IMAGE DATABASE
VIDEO ANALYSIS SECTION
VIDEO ANALYSIS RESULT
IMAGE SELECTION SECTION
LINK SECTION
DISPLAY SECTION
USER PROFILE DATABASE

サーバ/クライアント処理の一構成(2)

AN EXAMPLE OF ARRANGEMENT OF SERVER/CLIENT PROCESSING (2)

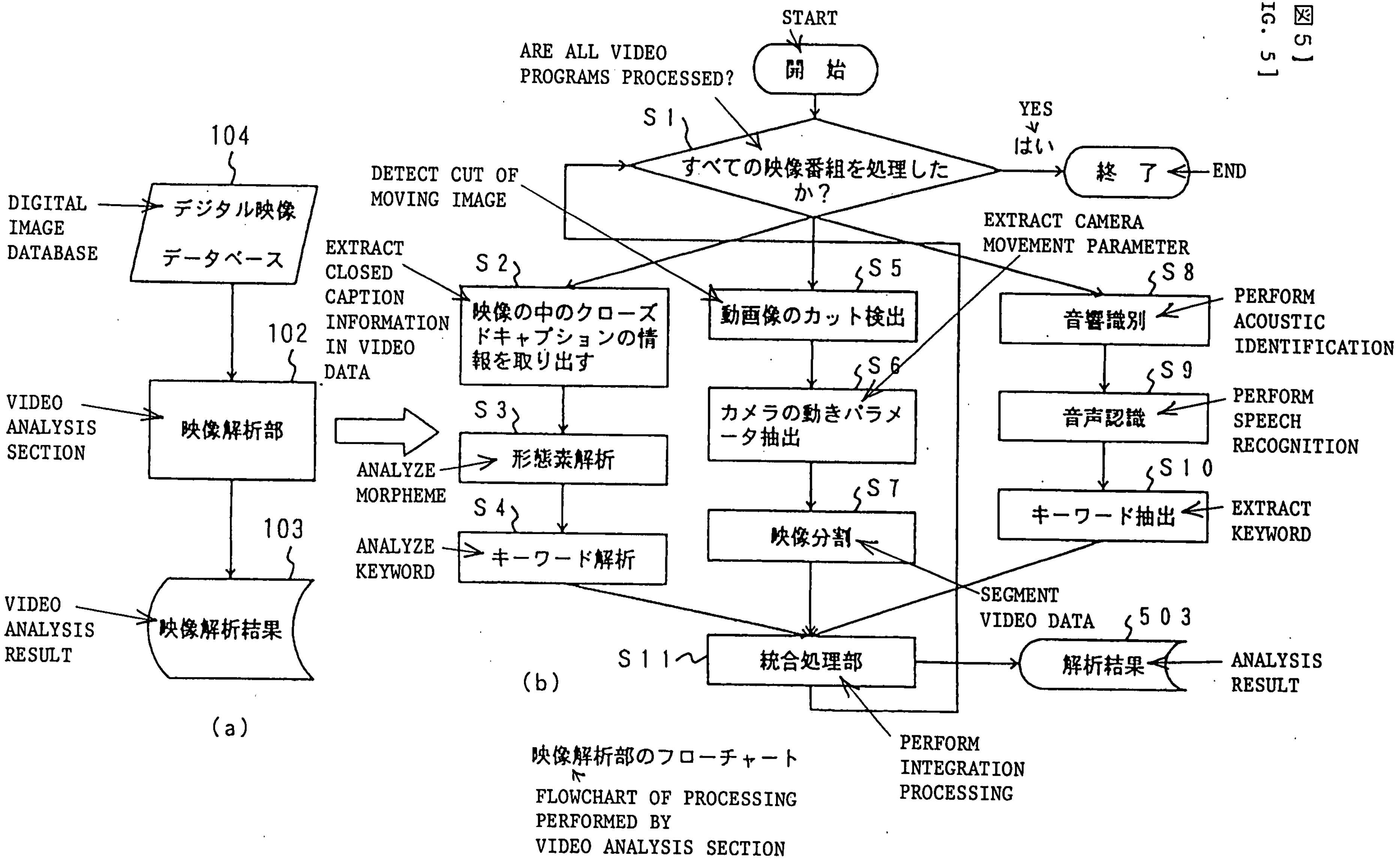
【図4】
[FIG. 4]


SERVER
サーバ
CLIENT
クライアント
101
デジタル映像データベース
DIGITAL IMAGE DATABASE
102
映像解析部
VIDEO ANALYSIS SECTION
103
映像解析結果
VIDEO ANALYSIS RESULT
104
個人プロファイル
USER PROFILE DATABASE
105
映像選択部
IMAGE SELECTION SECTION
106
映像関連付け部
LINK SECTION
107
表示部
DISPLAY SECTION

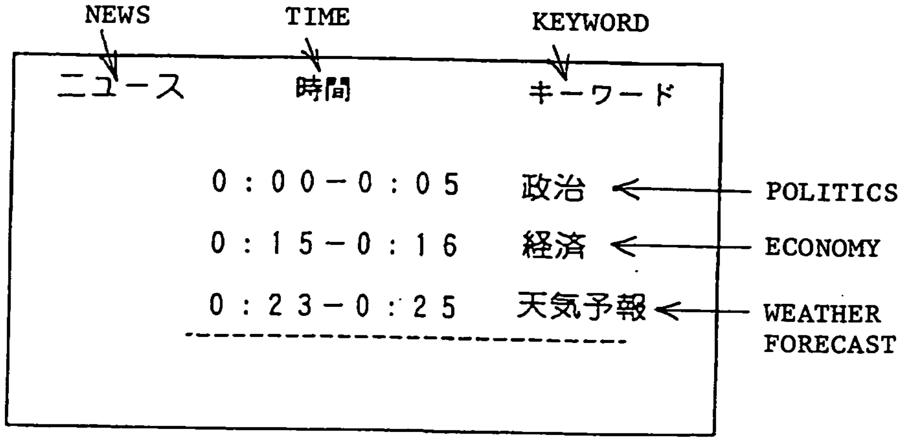
サーバ/クライアント処理の一構成(3)

AN EXAMPLE OF ARRANGEMENT OF SERVER/CLIENT PROCESSING (3)

【図5】
[FIG. 5]


START
開 始
ARE ALL VIDEO PROGRAMS PROCESSED?
S 1
すべての映像番組を処理したか？
YES
はい
終 了
END
104
DIGITAL IMAGE DATABASE
デジタル映像データベース
102
VIDEO ANALYSIS SECTION
映像解析部
103
VIDEO ANALYSIS RESULT
映像解析結果
(a)
DETECT CUT OF MOVING IMAGE
EXTRACT CLOSED CAPTION INFORMATION IN VIDEO DATA
S 2
映像の中のクローズドキャプションの情報を取り出す
S 3
ANALYZE MORPHEME
形態素解析
S 4
ANALYZE KEYWORD
キーワード解析
S 5
動画像のカット検出
EXTRACT CAMERA MOVEMENT PARAMETER
S 6
カメラの動きパラメータ抽出
S 7
映像分割
SEGMENT VIDEO DATA
S 8
音響識別
PERFORM ACOUSTIC IDENTIFICATION
S 9
音声認識
PERFORM SPEECH RECOGNITION
S 1 0
キーワード抽出
EXTRACT KEYWORD
S 1 1
統合処理部
PERFORM INTEGRATION PROCESSING
5 0 3
解析結果
ANALYSIS RESULT
(b)
映像解析部のフローチャート
FLOWCHART OF PROCESSING PERFORMED BY VIDEO ANALYSIS SECTION

【図6】
[FIG. 6]

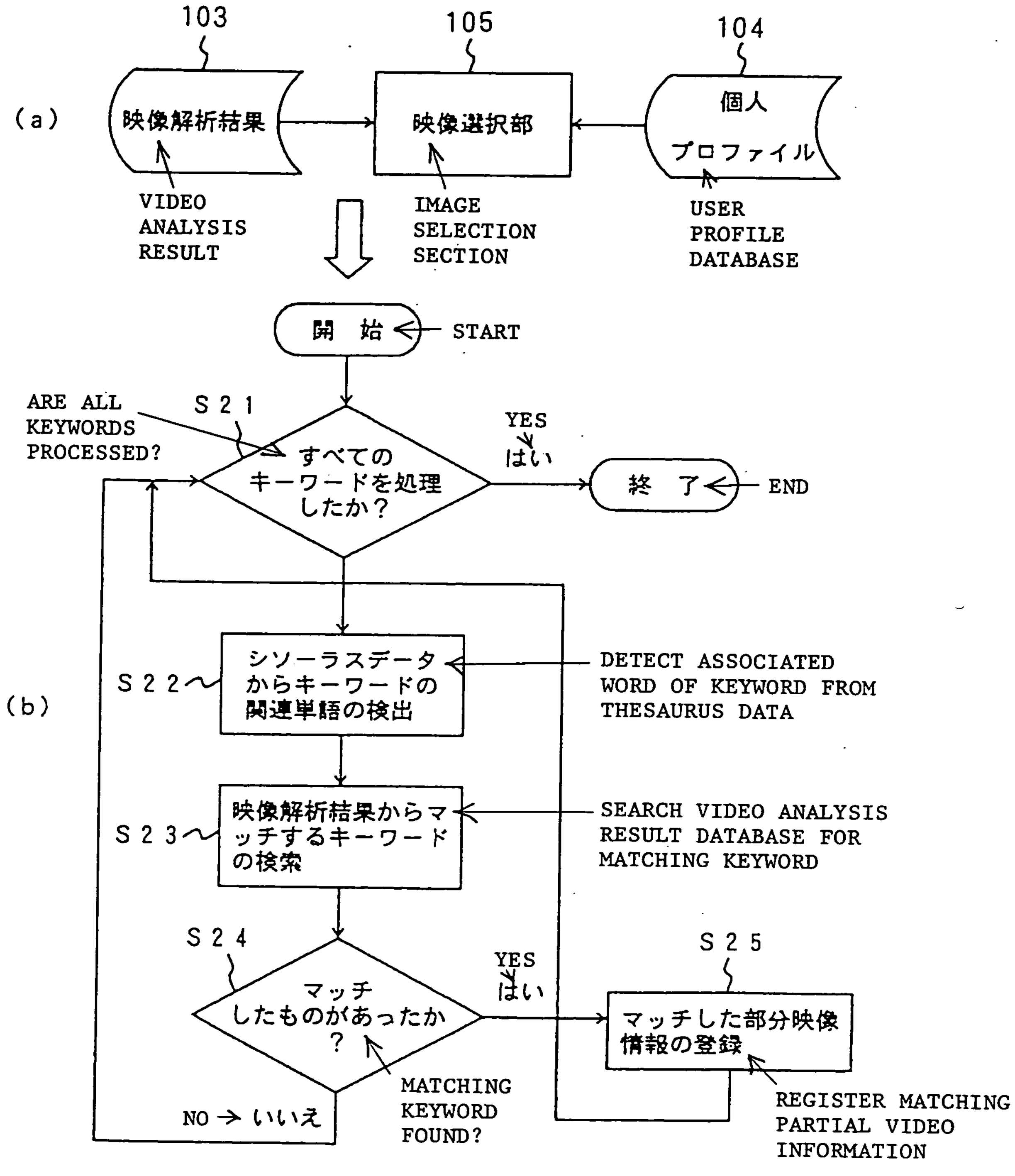

NEWS
TIME
KEYWORD
POLITICS
ECONOMY
WEATHER FORECAST

ニュース	時間	キーワード
	0：00－0：05	政治
	0：15－0：16	経済
	0：23－0：25	天気予報

解析結果の例
EXAMPLE OF ANALYSIS RESULT

【図７】

[FIG. 7]

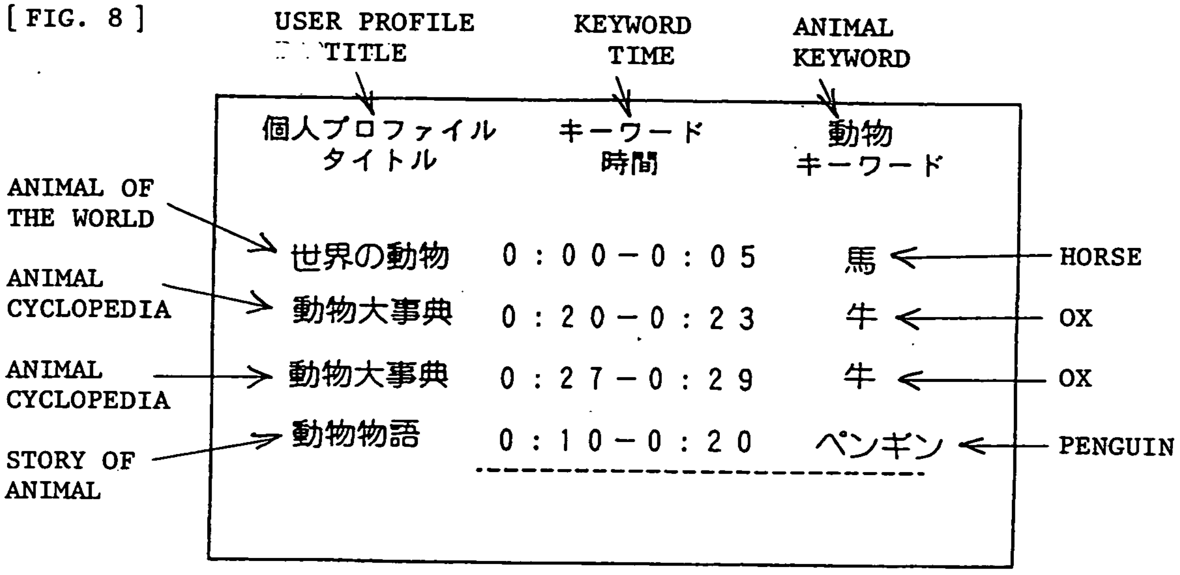

(a)
103
映像解析結果
VIDEO ANALYSIS RESULT
105
映像選択部
IMAGE SELECTION SECTION
104
個人プロファイル
USER PROFILE DATABASE
(b)
開始
START
S21
すべてのキーワードを処理したか？
ARE ALL KEYWORDS PROCESSED?
YES
はい
終了
END
S22
シソーラスデータからキーワードの関連単語の検出
DETECT ASSOCIATED WORD OF KEYWORD FROM THESAURUS DATA
S23
映像解析結果からマッチするキーワードの検索
SEARCH VIDEO ANALYSIS RESULT DATABASE FOR MATCHING KEYWORD
S24
マッチしたものがあったか？
MATCHING KEYWORD FOUND?
YES
はい
NO
いいえ
S25
マッチした部分映像情報の登録
REGISTER MATCHING PARTIAL VIDEO INFORMATION

映像選択処理の処理フロー

PROCESSING FLOW OF IMAGE SELECTION PROCESSING

【図8】

[FIG. 8]



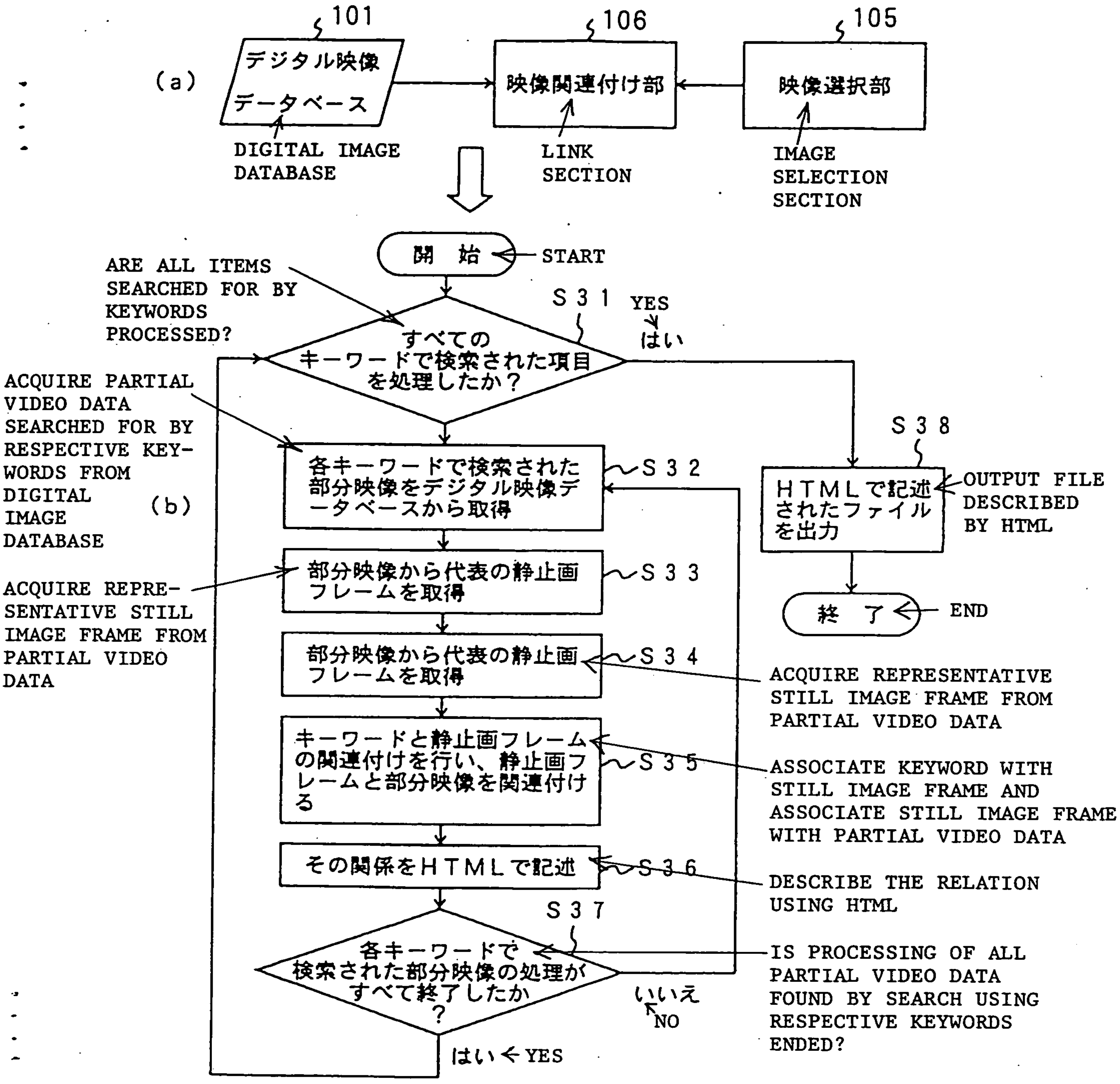
個人プロファイルタイトル	キーワード時間	動物キーワード
世界の動物	0：00－0：05	馬
動物大事典	0：20－0：23	牛
動物大事典	0：27－0：29	牛
動物物語	0：10－0：20	ペンギン

USER PROFILE TITLE
KEYWORD TIME
ANIMAL KEYWORD
ANIMAL OF THE WORLD
ANIMAL CYCLOPEDIA
ANIMAL CYCLOPEDIA
STORY OF ANIMAL
HORSE
OX
OX
PENGUIN

キーワードマッチング結果の記録

RECORD OF KEYWORD MATCHING RESULT

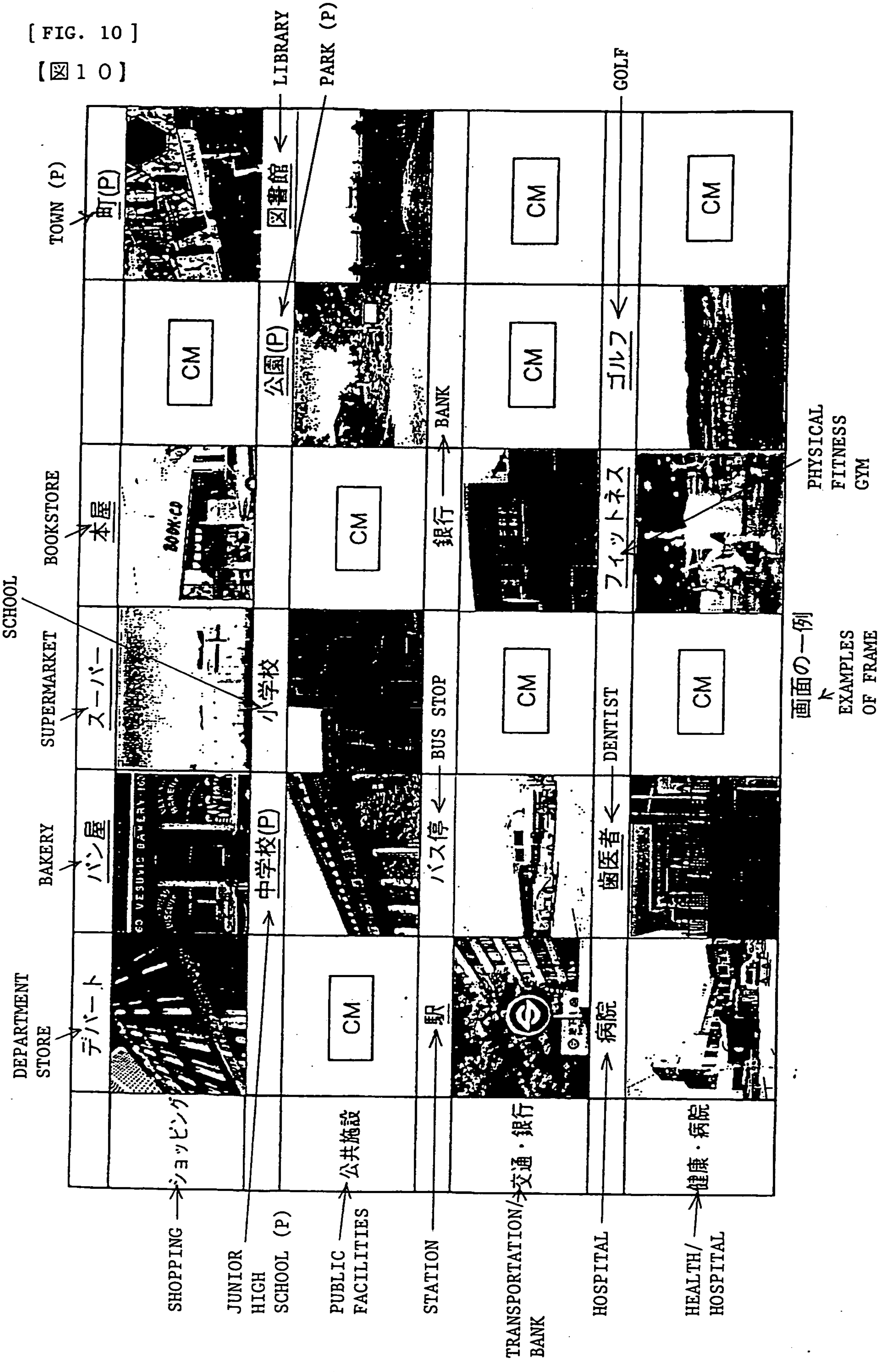
【図９】
[FIG. 9]


101
106
105
(a)
デジタル映像データベース
DIGITAL IMAGE DATABASE
映像関連付け部
LINK SECTION
映像選択部
IMAGE SELECTION SECTION
(b)
開　始
START
すべてのキーワードで検索された項目を処理したか？
ARE ALL ITEMS SEARCHED FOR BY KEYWORDS PROCESSED?
S31
はい
YES
各キーワードで検索された部分映像をデジタル映像データベースから取得
ACQUIRE PARTIAL VIDEO DATA SEARCHED FOR BY RESPECTIVE KEY-WORDS FROM DIGITAL IMAGE DATABASE
S32
部分映像から代表の静止画フレームを取得
ACQUIRE REPRE-SENTATIVE STILL IMAGE FRAME FROM PARTIAL VIDEO DATA
S33
部分映像から代表の静止画フレームを取得
ACQUIRE REPRESENTATIVE STILL IMAGE FRAME FROM PARTIAL VIDEO DATA
S34
キーワードと静止画フレームの関連付けを行い、静止画フレームと部分映像を関連付ける
ASSOCIATE KEYWORD WITH STILL IMAGE FRAME AND ASSOCIATE STILL IMAGE FRAME WITH PARTIAL VIDEO DATA
S35
その関係をＨＴＭＬで記述
DESCRIBE THE RELATION USING HTML
S36
各キーワードで検索された部分映像の処理がすべて終了したか？
IS PROCESSING OF ALL PARTIAL VIDEO DATA FOUND BY SEARCH USING RESPECTIVE KEYWORDS ENDED?
S37
いいえ
NO
はい
YES
S38
ＨＴＭＬで記述されたファイルを出力
OUTPUT FILE DESCRIBED BY HTML
終　了
END

映像関連付け部の処理フローを示す図
VIEW SHOWING PROCESSING FLOW PERFORMED BY LINK SECTION

[FIG. 10]

【図１０】


ショッピング
公共施設
交通・銀行
健康・病院
デパート
パン屋
スーパー
本屋
CM
町(P)
中学校(P)
小学校
公園(P)
図書館
駅
バス停
銀行
病院
歯医者
フィットネス
ゴルフ
画面の一例
SHOPPING
JUNIOR HIGH SCHOOL (P)
PUBLIC FACILITIES
STATION
TRANSPORTATION/BANK
HOSPITAL
HEALTH/HOSPITAL
DEPARTMENT STORE
BAKERY
SUPERMARKET
ELEMENTARY SCHOOL
BOOKSTORE
TOWN (P)
LIBRARY
PARK (P)
BUS STOP
BANK
GOLF
DENTIST
PHYSICAL FITNESS GYM
EXAMPLES OF FRAME

BEST AVAILABLE COPY

[Document] ABSTRACT

[Abstract]

[Object] An object of the present invention is to make it possible to automatically select a certain portion of video data required by a user from various programs as an object of recording or watching and listening to, in consideration of a problem that all the programs have to be selected and recorded or watched and listened to, in spite of the user's interest in only a portion of the video data.

[Means for Achieving the Object] An image providing apparatus has means (102) for obtaining information representing the contents of video data by subjecting the video data to analysis (moving image analysis, acoustic/speech analysis, and text analysis), and adding an index to the video data on the basis of the obtained information. The image providing apparatus selects a plurality of partial video data items from one or a plurality of video data by means of selecting means (105) on the basis of that index and personal profile data registered in advance or profile information input on-line or inquiry information. The selected video data items are associated with each other by associating means (106) so as to be displayed and easily watched, thereby providing a user with only video data required by the user.

[Elected Figure] FIG. 1